



The Battle of Neighborhoods

-The New York City perspective –

Francesco Morena for Coursera Capstone Project

Predicting optimum location is valuable for new Restaurant business

New York City is the financial capital of USA. It is multicultural, provides lot of business opportunities, it is business friendly environment and attracted many different players into the market.

- Business Problem:
 - Choise of first neighborhood to start restaurant business
 - Choise of neighborhood that has good amount of customers
 - Easy to replicate
 - Low competition
- Succes Criteria:
 - Best neighborhood which meets aforementioned criteria.

Questions that can be asked:

- What is best location in **Brooklyn** that has potential **Italian Restaurant market**?
- Which areas are saturated with Italian restaurants, which untapped?

Data acquisition

To solve the aim of the project, we will need the following data:

- List of Neighborhoods in NYC, USA:
 - Data source : https://cocl.us/new_york_dataset
 - Description : This data set contains the required information. And we will use this data set to explore various neighborhoods of new york city.
- Venue data, particularly related to Italian Restaurant:
 - Data source : Fousquare API
 - Description : By using this api we will get all the venues in each neighborhood. We can filter these venues to get only italian restaurants.

Methodology

In general, this project would be encompassing a series of Data Science techniques, including, but not limited to, Data Cleaning, Data Wrangling and Machine Learning (K-Means clustering algorithm)

- Approach:
 - Collect the New York City data from https://cocl.us/new_york_dataset
 - Using FourSquare API we will find all venues for each neighborhood
 - Filter out all neighborhoods of Brooklyn
 - Filter out all venues that are Italian Restaurants, Office, University, Hischool
 - Analyze venues of interest using FouSquare API
 - Using rating for each neighborhoods , we will sort that data
 - Visualize the Ranking of neighborhoods using folium library(python)
 - K-Means clustering and Clustering Analisys

Methodology

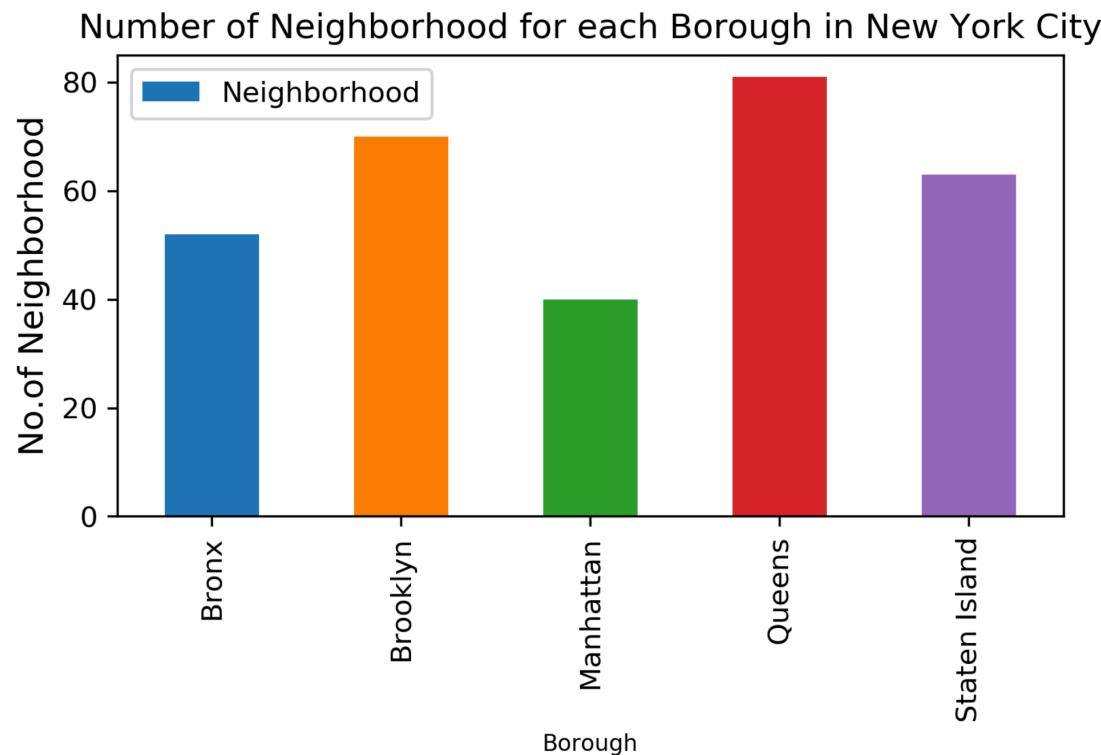
For each neighborhood, the sums of the office, school, university and italian restaurant were computed and for each of this 4 categories, a weight (or penalty) has been defined according to what stockholder considers the most important.

- Italian restaurant have been weighted with -1, since stockholder wants to avoid concurrence
- HighSchools have been weighted with 1, since employees/student are good customers
- Universities have been weighted with 1.5, since employees/students are good customers
- Offices have been weighted with 2, since employees are even better customers
- A score was computed for each locality as the weighted sum of the number of venues in each of the 4 categories (school, university, office, Italian Restaurant)
- Lastly, K-Means clustering is performed on this data set to return clusters, or categories of neighborhoods in terms of number of Italian Restaurant

Questions that can be asked using the aforementioned datasets:

- What is best location in Brooklyn that has potential Italian Restaurant market?
- Which areas are saturated with Italian restaurants, which untapped?

There are 5 Borough and 306 Neighborhood across New York City

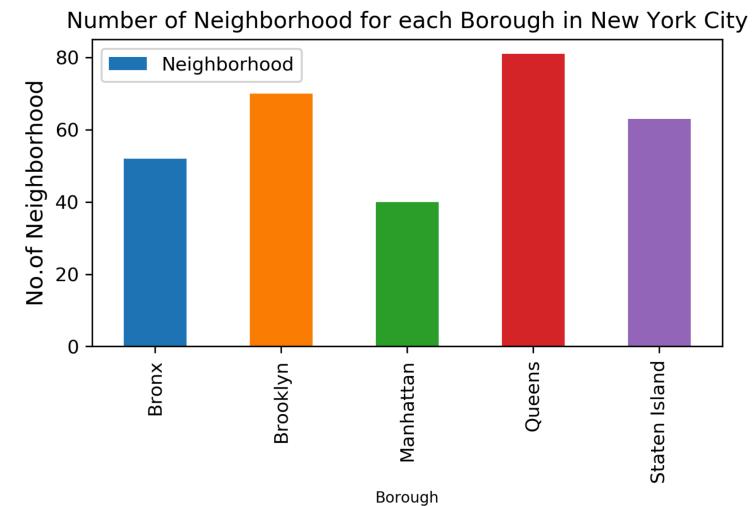
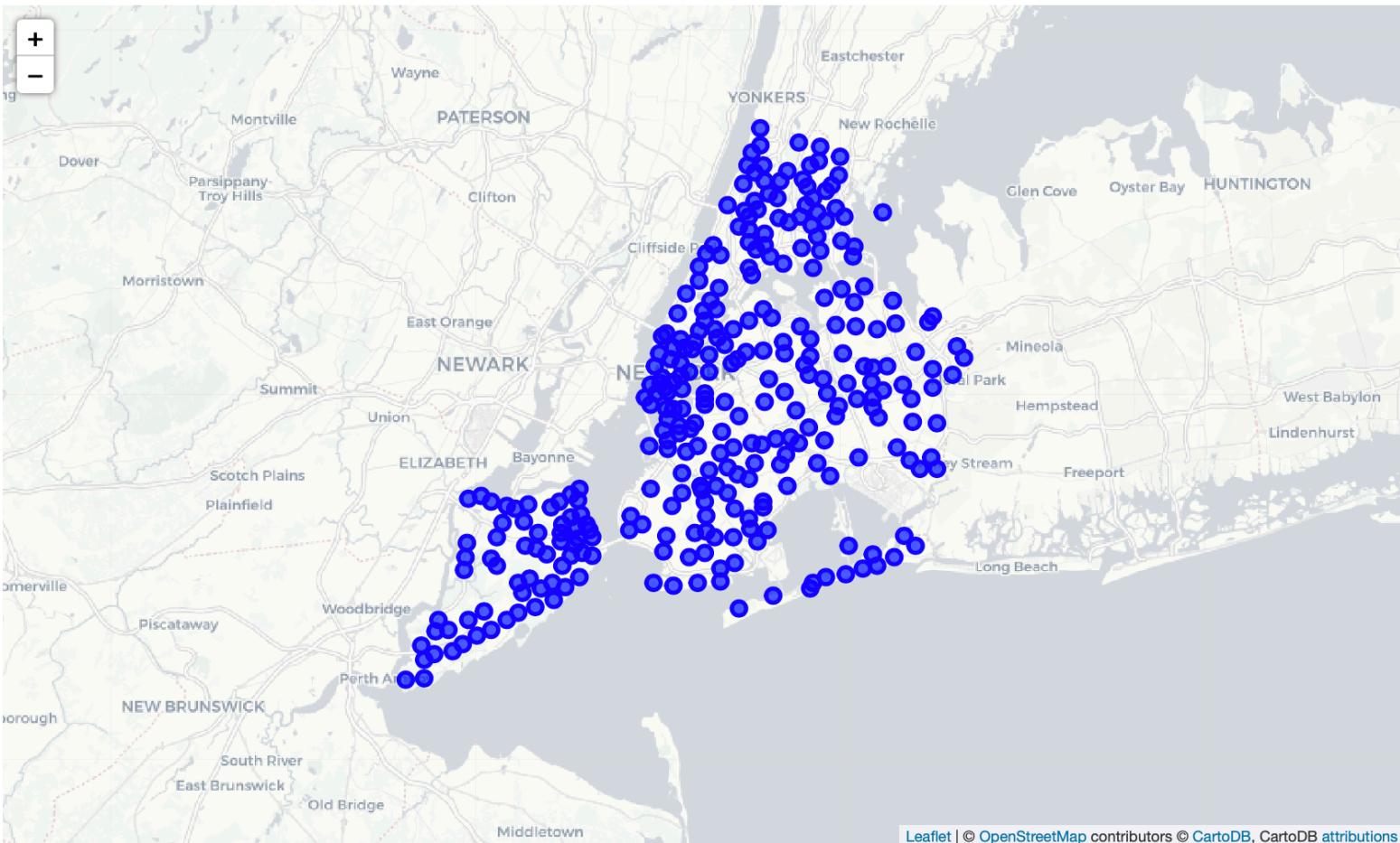


Queens has the highest number of neighborhoods, followed by Brooklyn and Staten Island

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

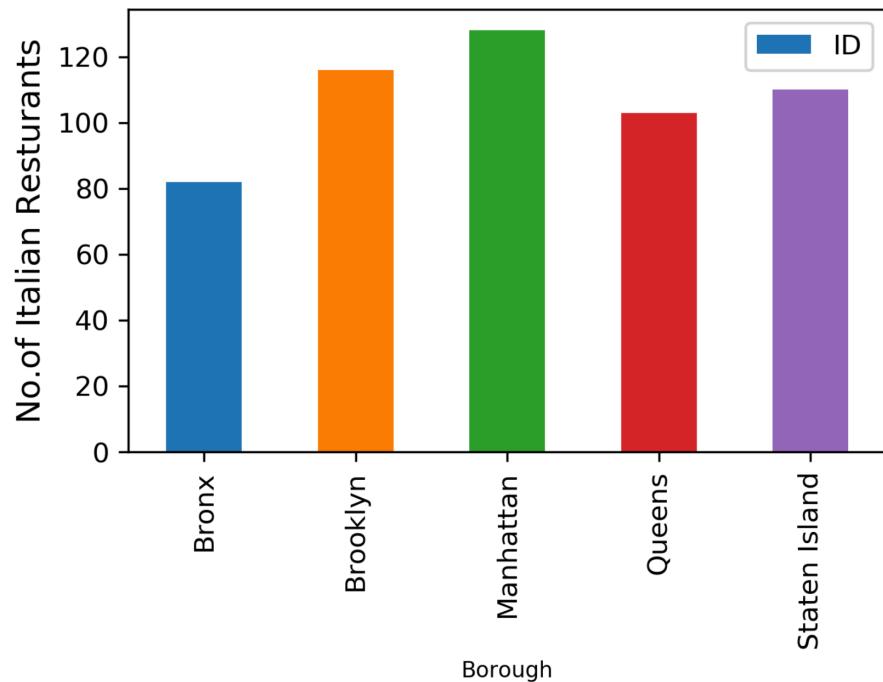
df.shape
(306, 4)

Map of 306 Neighborhood across New York City



The are 539 Italian Restaurants across Borough in New York City

Number of Italian Restaurants for each Borough in New York City



```
italian_rest_ny.head()
```

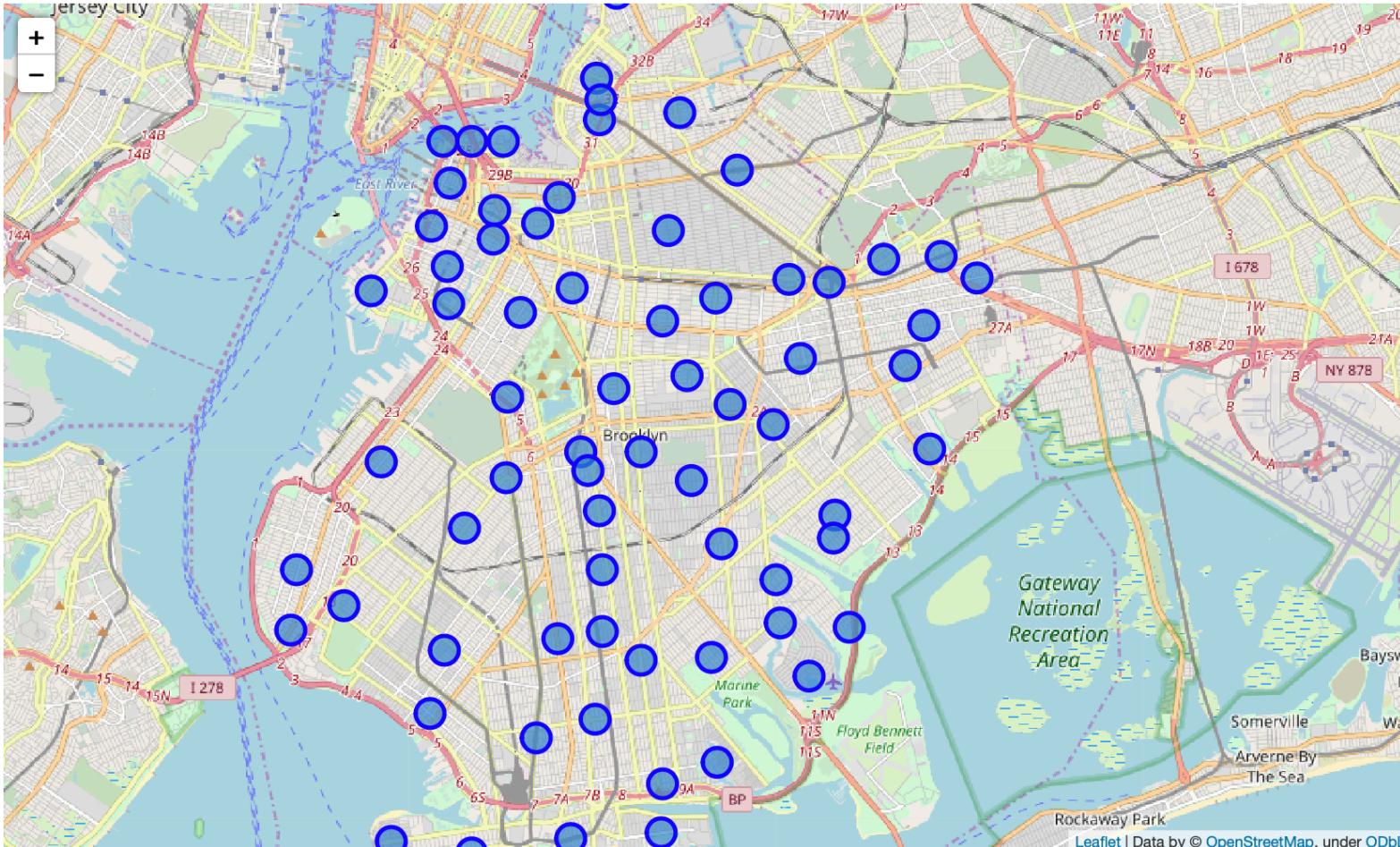
	Borough	Neighborhood	ID	Name
0	Bronx	Riverdale	55aaeee4d498e3cbb70e625d6	Bella Notte Pizzeria
1	Bronx	Kingsbridge	55aaeee4d498e3cbb70e625d6	Bella Notte Pizzeria
2	Bronx	Woodlawn	511edb6de4b0d58346fd272d	Patrizia's Of Woodlawn
3	Bronx	Woodlawn	4d3cb3026b3d236a066a6364	Rivers Edge
4	Bronx	Baychester	4c9518076b35a143d5dc21dc	Fratelli's

```
italian_rest_ny.shape
```

(539, 4)

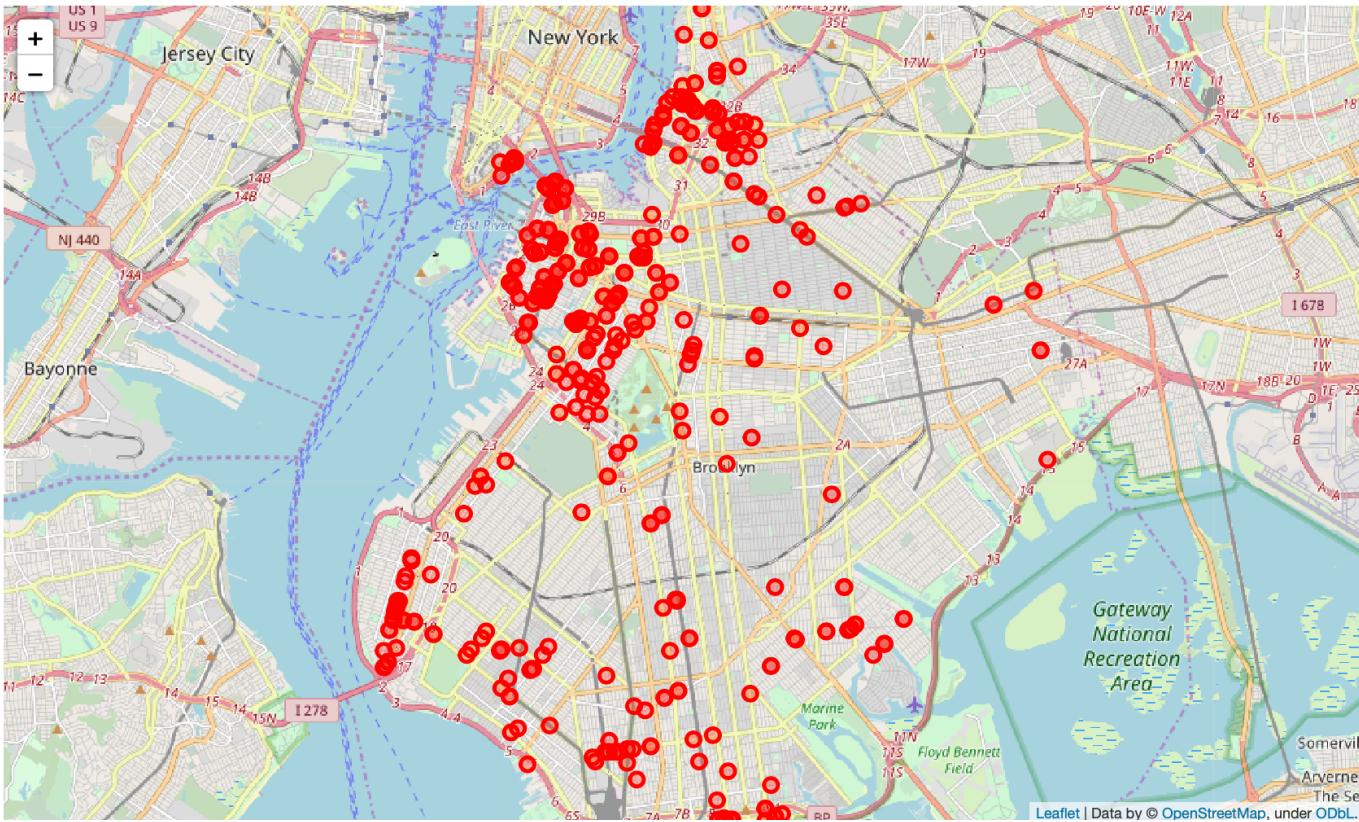
Manhattan has the largest number of Italian restaurants, followed by Brooklyn. Bronx has the smallest number of Italian restaurants.

Map of 70 Neighborhoods in Brooklyn



Due to the aim of the project and the request of the stockholder, we taking into account Italian Restaurants in a specific area: Brooklyn

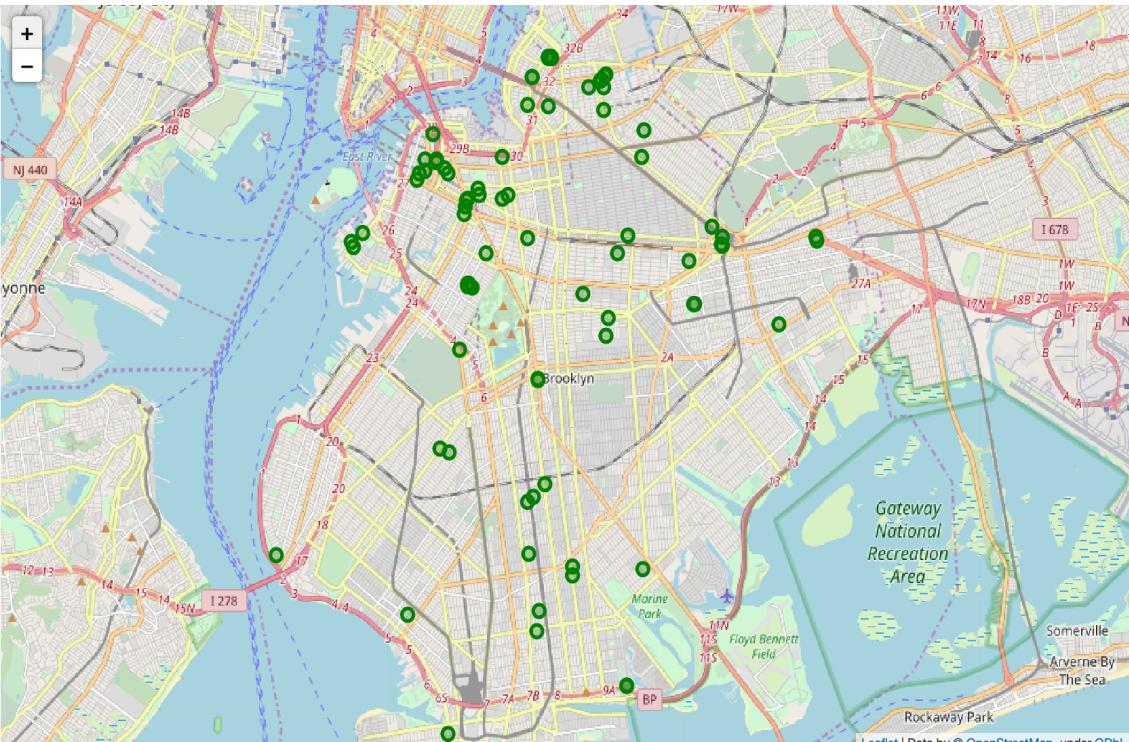
Map of Italian Restaurants in Brooklyn



Count
Italian Restaurant 536
Bar 3
Gastropub 2
Food & Drink Shop 2
Pizza Place 1

The are 536 Italian Restaurants across Brooklyn

Map of HighSchools in Brooklyn



The are 81 HighSchools across Brooklyn

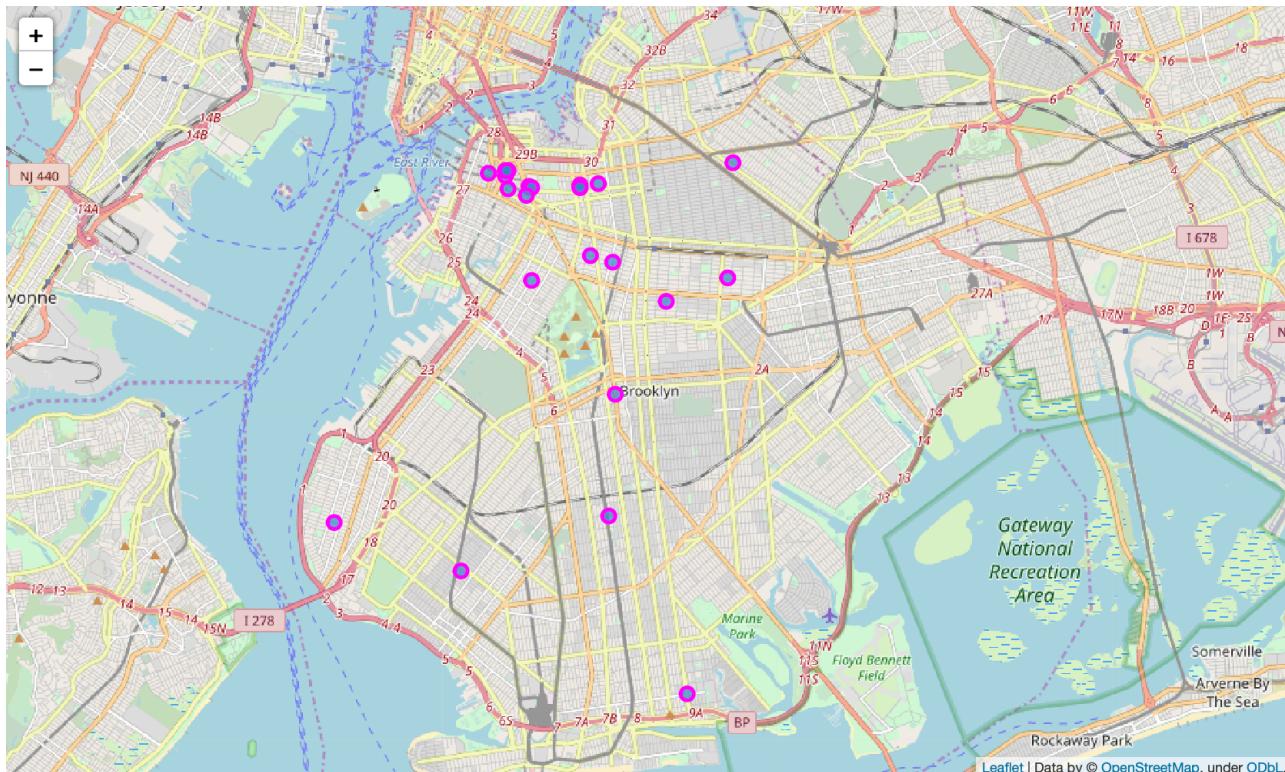
`broo_venues_highschools.head()`

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue
0	Manhattan Terrace	40.614433	-73.957438	Imagine Aca
1	Crown Heights	40.670829	-73.943291	St. Marks S
2	Windsor Terrace	40.656946	-73.980073	Bishop Ford Central Catholic High S
3	Prospect Heights	40.676822	-73.964859	Acorn Community High S
4	Brownsville	40.663950	-73.910235	Teachers Preparatory High S

`broo_venues_highschools.shape`

(81, 7)

Map of Universities in Brooklyn



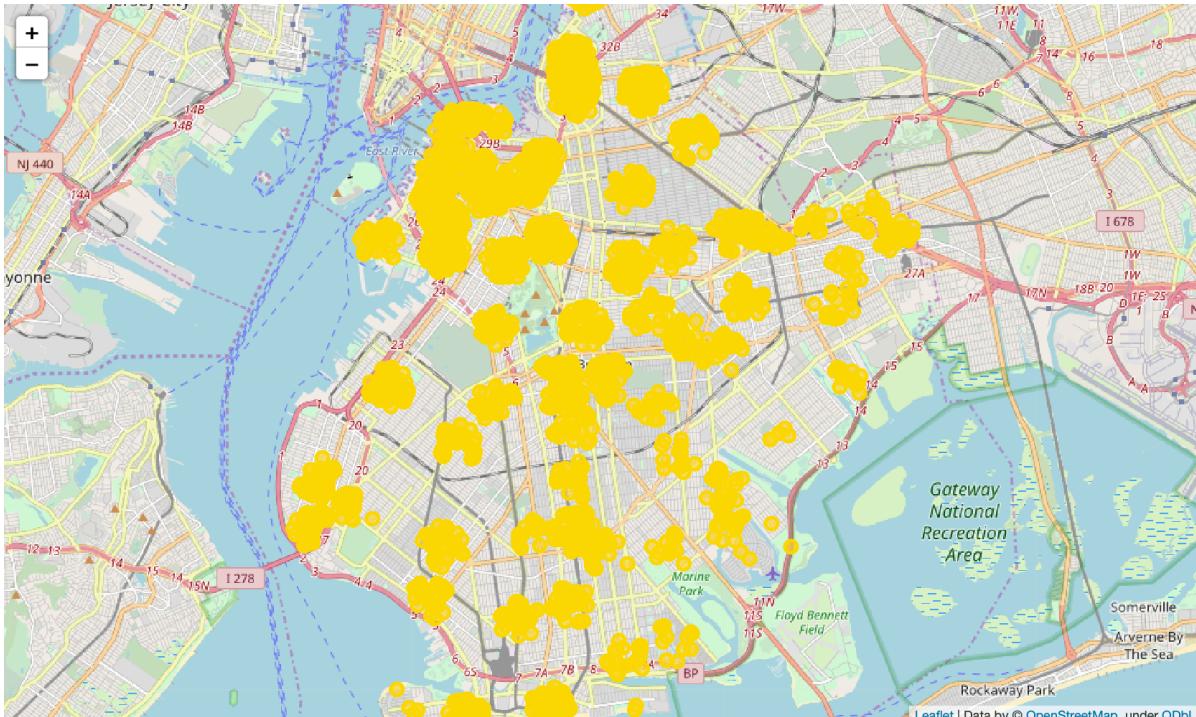
The are 24 Universities across Brooklyn

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	
0	Bay Ridge	40.625801	-74.030621	Stepping Sto
1	Bensonhurst	40.611009	-73.995180	In
2	Sheepshead Bay	40.586890	-73.943186	Financi
3	Crown Heights	40.670829	-73.943291	Cornell University Cooperative
4	Prospect Heights	40.676822	-73.964859	Bacca

`broo_venues_uni.shape`

(24, 7)

Map of Offices in Brooklyn



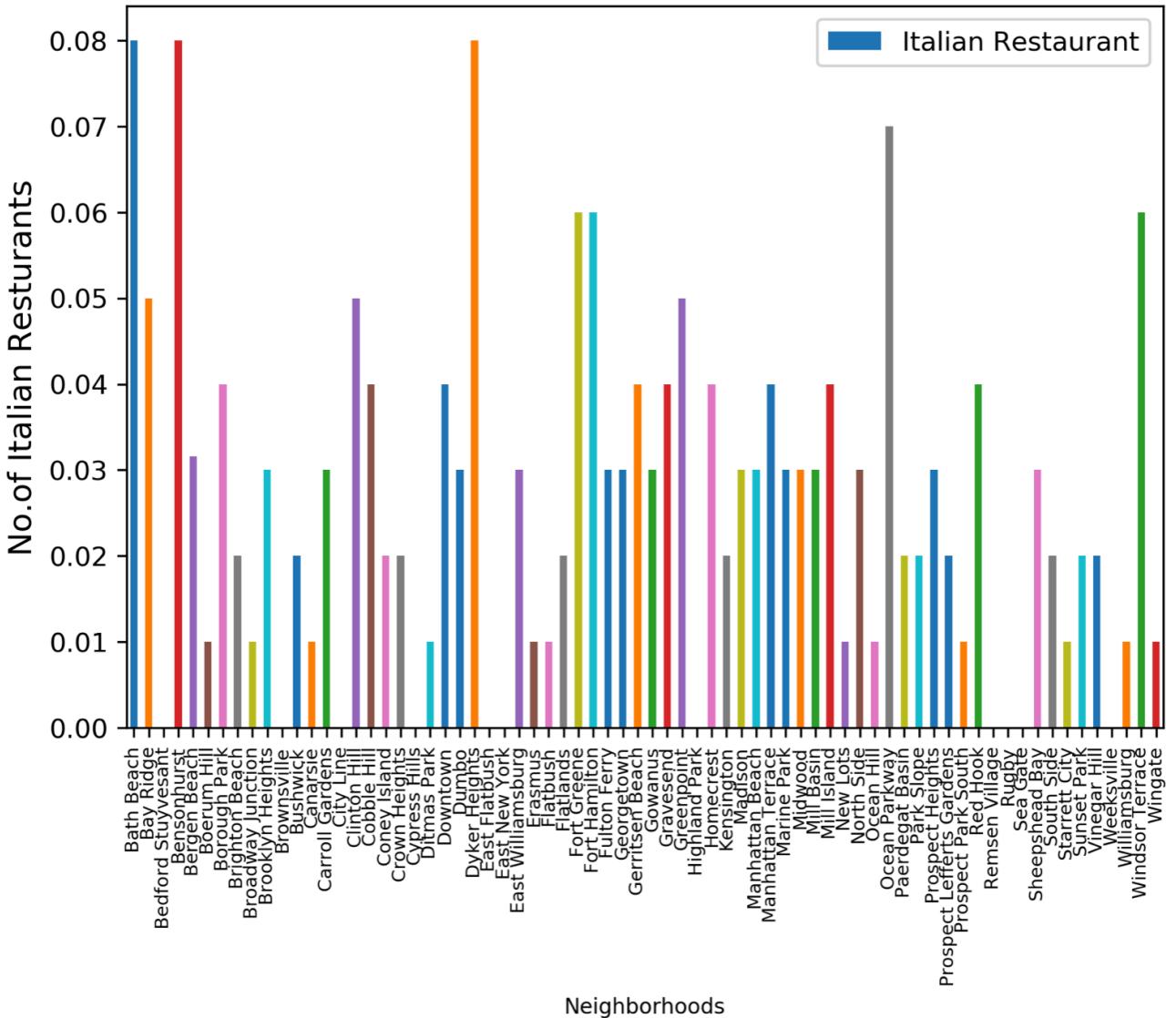
	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	VenueName
0	Bay Ridge	40.625801	-74.030621	WishWash Carpet Cleaning & Restoration of Bay ...
1	Bay Ridge	40.625801	-74.030621	R&G Brenner Income Tax
2	Bay Ridge	40.625801	-74.030621	Integrated Financial Services
3	Bay Ridge	40.625801	-74.030621	Polizzi Fine Art
4	Bay Ridge	40.625801	-74.030621	H And R Block

broo_venues_office.shape

(3278, 7)

The are 3278 Offices across Brooklyn

Frequencies Plot of Italian Restaurants for each Neighborhood in Brooklyn



Frequencies Plot

Results

	Neighborhood	Score
86	Downtown	195.5
283	Dumbo	193.0
49	Greenpoint	191.0
64	Brooklyn Heights	184.5
70	Park Slope	181.5
96	North Side	176.0
87	Boerum Hill	168.5
280	Vinegar Hill	165.0
99	Fort Hamilton	163.0
53	Manhattan Terrace	160.0
61	Williamsburg	154.0

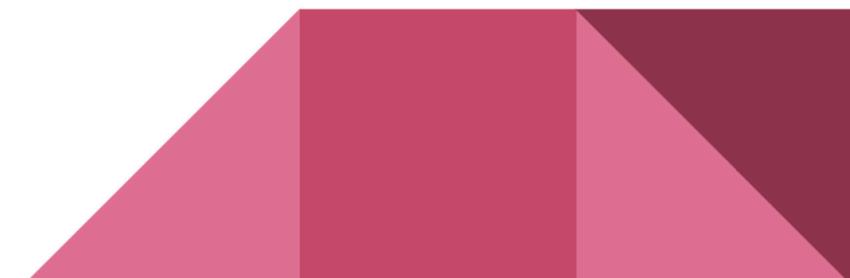
The neighborhoods with the best score are:

Downtown, Dumbo, Greenpoint, Brooklyn Heights and Park Slope are some of the best neighborhoods for stockholder in order to open his Italian Restaurant.

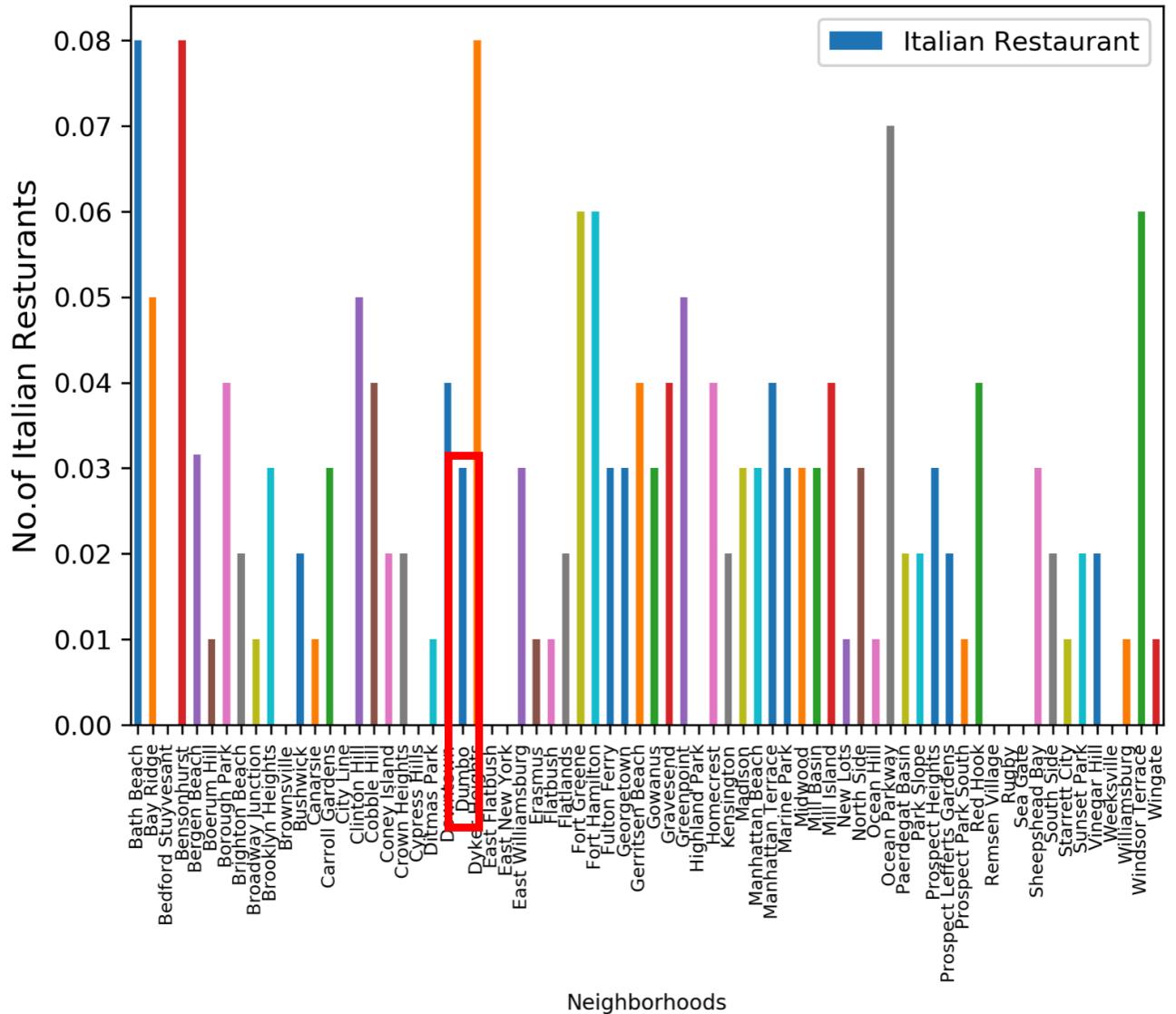
These options maximize the number of potential customers from Offices and Universities and at the same time have not too large competitor.

However, analyzing the Neighborhood with the best score, Dumbo would seem to be the most suitable as the one with the lowest number of Italian restaurants (7), compared to the Downtown (17), Greenpoint (9), Brooklyn Heights (22) and Park Slope (26).

Dumbo with a score of 193.0 being
the best option

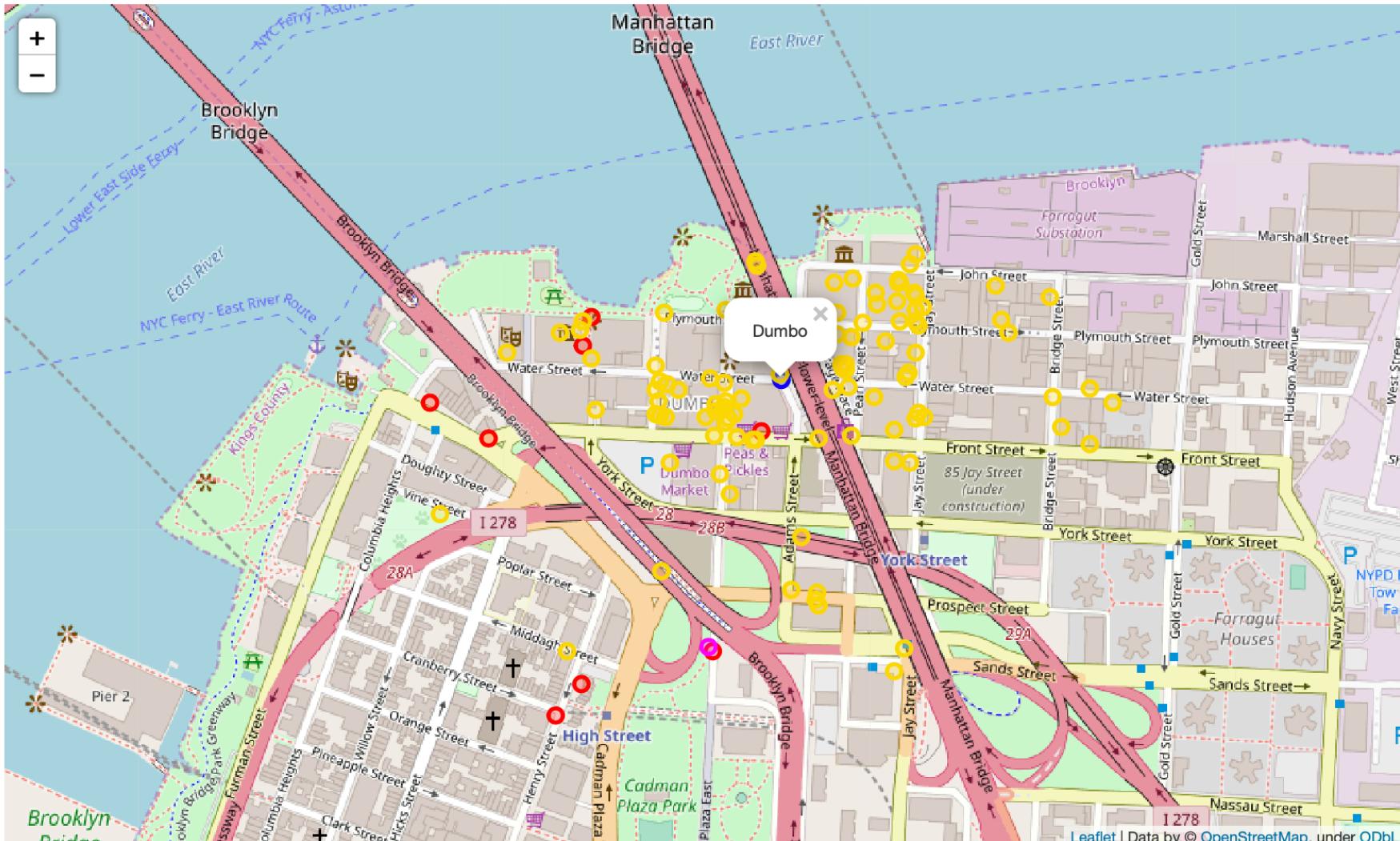


Frequencies Plot of Italian Restaurants for each Neighborhood in Brooklyn



Frequencies Plot
with Dumbo
highlighted

Best Place for the Italian Restaurant in Brooklyn is 'DUMBO'



K-Means Clustering

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
count	10.0	NaN	NaN	NaN	NaN
freq	NaN	50.00%	20.00%	20.00%	20.00%
mode	NaN	Pizza Place	Italian Restaurant	Beach	Bakery

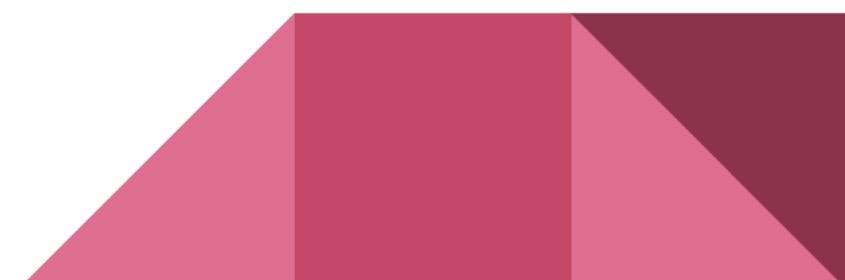
Top 5 most common venues in cluster 0

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
count	10.0	NaN	NaN	NaN	NaN
freq	NaN	40.00%	30.00%	20.00%	30.00%
mode	NaN	Caribbean Restaurant	Pizza Place	Café	Coffee Shop

Top 5 most common venues in cluster 1

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
count	7.0	NaN	NaN	NaN	NaN
freq	NaN	100.00%	28.57%	28.57%	42.86%
mode	NaN	Caribbean Restaurant	Café	Donut Shop	Pizza Place

Top 5 most common venues in cluster 2



K-Means Clustering

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
count	11.0	NaN	NaN	NaN	NaN
freq	NaN	54.55%	27.27%	18.18%	36.36%
mode	NaN	Pizza Place	Bagel Shop	Pizza Place	Bagel Shop

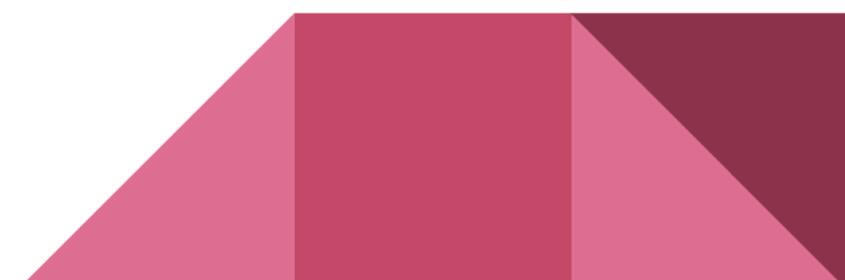
Top 5 most common venues in cluster 3

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
count	11.0	NaN	NaN	NaN	NaN
freq	NaN	72.73%	27.27%	36.36%	27.27%
mode	NaN	Pizza Place	Discount Store	Supermarket	Fast Food Restaurant

Top 5 most common venues in cluster 4

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
count	21.0	NaN	NaN	NaN	NaN
freq	NaN	23.81%	28.57%	23.81%	19.05%
mode	NaN	Bar	Pizza Place	Coffee Shop	Italian Restaurant

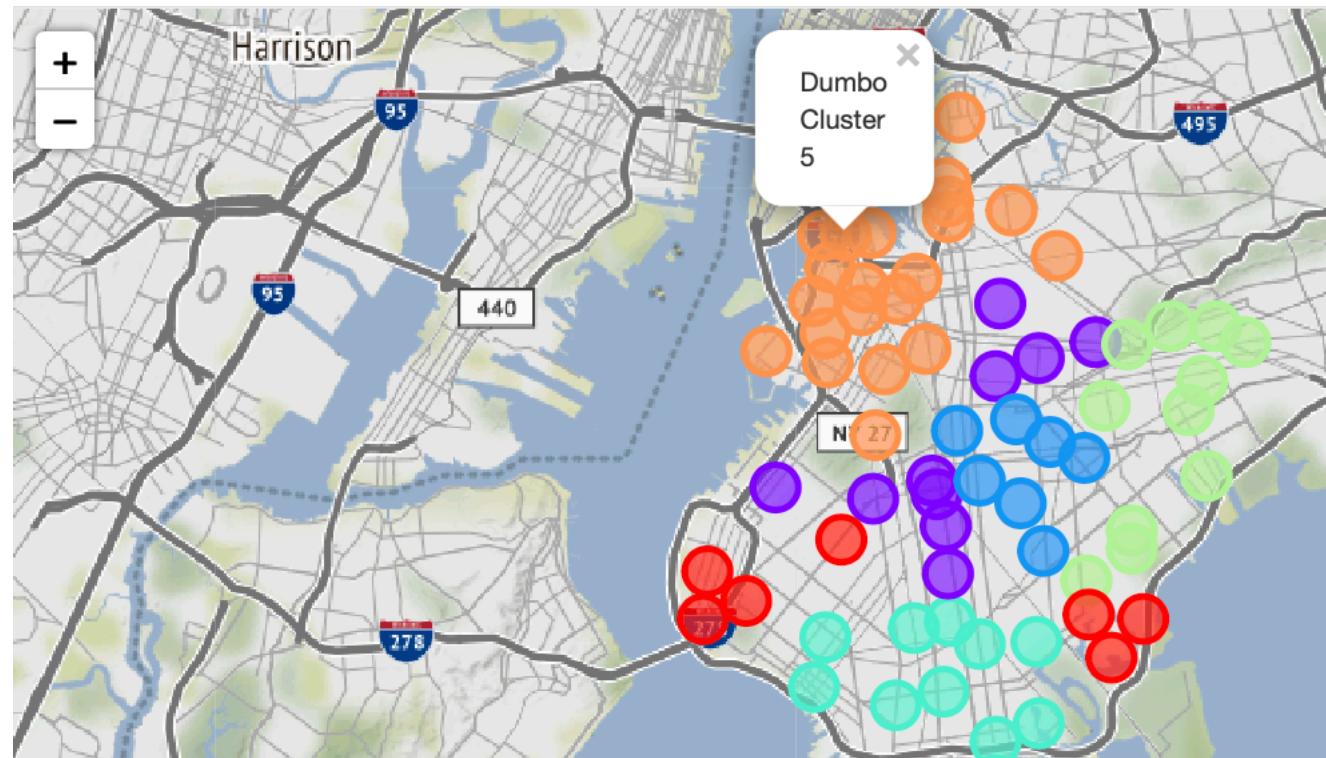
Top 5 most common venues in cluster 5



K-Means Clustering results

The following are the highlights of the 6 clusters:

- Pizza Place and Italian Restaurant are the most popular in the cluster 0. The Italian Restaurants are very popular in this cluster, especially in Dyker Heights, Fort Hamilton and Bay Ridge areas.
- In clusters 1 and 2 predominant is the Caribbean Restaurant, which suggests a prevalent Caribbean presence, followed by Cafe and Pizza Place;
- Pizza Place, Bagel shop, and Grocery Store are the most popular in the cluster 3.
- Pizza Place, Discount Store, Supermarket, and Fast Food Restaurant are the most popular in the cluster 4.
- Bar, Pizza Place, Coffee Shops are the most popular in the cluster 5.



The best choice to open an Italian Restaurant could be cluster 5.

In this cluster reside our choice Dumbo

Conclusion and future directions

Cluster 5 is the most viable clusters to create a brand Italian Restaurant. Their proximity to other amenities are paramount. In this cluster, even if there are other Italian restaurants, the market is not saturated. So the rivalry could be minimal.

In conclusion, this project would have had better results if there were more data in terms of crime data within the area, parking access, bus and metro stop, and allowance of more venues exploration with the Foursquare (limited venues for free calls).

Also, getting the ratings and feedbacks of the current restaurants within the clusters would have helped in providing more insight into the best location.